2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) | 978-1-6654-9079-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIBA56860.2023.10165349

# Building A Deep Learning Model for Multi-Label Classification of Natural Disasters

Qiang Cao[1], Yan Liu[2*], Guangxu Wang[1], Yuxin HE[1], Kuanglan Wang[1], Shaoyi (Stephen) Liao[1], Lixin Pu[3]

1.City University of Hong Kong Shenzhen Research Institute, Shenzhen, China
2.City University of Hong Kong, Hong Kong, China
3. Shenzhen Institute for Advanced Study, UESTC,Shenzhen 518038, China
qiangcao2-c@my.cityu.edu.hk, yliu627-c@my.cityu.edu.hk, guawang9-c@my.cityu.edu.hk, yuxinhe3-c@my.cityu.edu.hk, klwang4@cityu.edu.hk, issliao@cityu.edu.hk, plx@uestc.edu.cn

*Abstract*—**Natural disasters, such as earthquakes, hurricanes/typhoons and wildfires, usually cause severe damage. Disaster response and management is a great challenge to the authority. Current studies usually focus on a single disaster identification using social media data. In reality, there are relationships among different types of disasters. And several disasters may happen simultaneously. In this study, we explore the role of the deep learning model in multi-label disaster classification. We build a deep CNN model for multi-label classification with the instruction of a high-order strategy. We train and validate our model using a professional low-altitude disaster dataset, LADI . We find our proposed deep learning model with the transfer learning method outperforms many other machine learning models in the previous study.**

***Keywords—disaster classification; multi-label learning; deep learning***

## I. INTRODUCTION

Natural disasters, from earthquakes and hurricanes/typhoons to floods and wildfires, have the power not only to injure/kill residents but also to create a substantial expense for governments, residents, and businesses. In 2021, a natural disaster caused around 280 billion dollars in economic losses in the world (https://www.forbes.com/sites/joewalsh/2022/01/10/us-natural-disasters-cost-145-billion-in-2021---3rd-costliest-year-on-record/?sh=760c2a294606). Disaster response and management is a great challenge to the authority.

With the prevalence of social media, the public could quickly understand disaster situations. And some researchers have used social media data, such as images, and tweets, to build a machine learning model for disaster classification [1]–[3]. However, most user-generated contents on social media are short, ambiguous words with images taken by mobile phones. Due to the data quality problem of social media contents, the capabilities of recent machine learning models could not meet disaster response and management's needs, such as support for search and rescue [4]. Besides, current studies usually focus on a single disaster identification. In reality, there are relationships among different types of disasters. And several disasters may happen simultaneously. For example, flooding may happen after a hurricane. And fire/smoke could be found after an earthquake. Therefore, identifying multiple disasters at the same time is more useful to support disaster response and management compared to traditional single disaster identification.

To develop technology to support disaster response and management, some professional organizations, such as the Civil Air Patrol, have collected a large scale of images related to various disasters [5]. This dataset has three key distinctions: first, low altitude and oblique perspective; second, high quality taken by professional equipment; third, multi-label annotations. One image may have 1 to 5 disaster categories based on its content. Therefore, we aim to answer the following research questions:

(1) How to utilize the low-altitude, high-quality(resolution) image to help the disaster response and management?

(2) How to build an advanced deep learning model to help identify multiple disasters simultaneously?

In this study, we explore the role of the deep learning model in multi-label disaster classification. We build a deep CNN model for multi-label classification with the instruction of a high-order strategy [6]. We train and validate our model using a professional low-altitude disaster dataset, LADI. It contains many high-quality disaster images captured by professional teams. We find our proposed deep learning model with a transfer learning method outperforms many traditional machine learning models with a first-order strategy. Besides, we also find the proposed deep learning model outperforms other traditional machine learning models with the higher-order strategy. Finally, we also explore performs of different deep learning models. We find the difference between different deep learning models is quite small.

The remainder of this study is organized as follows. We review related works in the literature review section. The research framework, proposed new model, dataset, and evaluation metrics are introduced in the research method section. The next section reports the experiment results. Finally, we present the conclusions of the study, its limitations, and future research directions in the last section.

## II. LITERATURE REVIEW

### A. Deep Learning in Disaster Image Classification

Disaster response and management is a complex and information-intensive activity. Decision making for disaster response highly depends on the disaster type and severity. Since social media is very prevalent in recent years [7], using user-generated content from social media for disaster classification and severity estimation has been discussed in several studies [1], [4]. Some studies use images posted by individuals to train and test the deep learning model for disaster classification [8], [9]. Convolutional neural network (CNN) is one of the widely used deep learning models in image classification. To further improve the classification performance, some researchers propose using multi-modal data for disaster classification. And integrating text with the image is well studied in recent years [2], [4]. Although those studies have made great progress in disaster classification, there are still several drawbacks: 1) the quality of social media data is low 2) Most studies focus on single-label disaster classification.

To further improve the performance of the deep learning model in disaster classification, some studies build a new high-quality dataset for model training and testing. Remote sensing data is used for sea ice classification using a residual network [10]. The satellite image is also used for building damage assessment [11]. Although various high-quality data collected by advanced equipment help disaster classification, most relevant studies only focus on a single type of disaster classification, which limits the application and extension of deep learning models.

### B. Deep Learning in Multi-label Image Classification

Multi-label learning focuses on the problem that a single instance is associated with a set of labels simultaneously [6]. For example, a patient may be diagnosed with "high blood pressure", "diabetes", and "stroke" simultaneously. In the context of a natural disaster, an image may show the scene of a hurricane, flood, and road washout simultaneously. There are some studies applying the deep learning method to multi-label classification in the healthcare field. A study develops a deep learning model to simultaneously diagnose three types of pelvic organ prolapse using medical images [12]. Another study builds three CNN models to automatically detect and classify the lesions of diabetic retinopathy using medical images [13]. Some researchers also use the DenseNet to detect multiple mutually non-exclusive chest diseases[14]. Although there are a lot of studies in the healthcare field that focus on multi-label classification, using deep learning methods for multi-label disaster classification very few. One study proposes a CNN model for multi-label classification[15]. A real hurricane dataset is used for storyline generation.

Since social media data limits the performance of deep learning, we would like to build a new deep learning model to process high-quality (resolution) images captured by a professional organization. And then, we use transfer learning to further improve the performance of our deep learning model. Finally, the multi-label classification deep learning model could detect several different natural disasters simultaneously.

## III. RESEARCH METHOD

### A. Multi-label Classification

Multi-label learning focuses on the problem that a single instance is associated with a set of labels simultaneously [6]. There are three types of strategies to cope with the multi-label problem from the perspective of label correlations:

First-order strategy: Decomposing multiple labels into several independent binary labels without considering the relationship among different labels.

Second-order strategy: Considering pairwise relations between labels, such as fire and not fire.

High-order strategy: Imposing all other labels' influences on each label.

To utilize the correlation between different labels, we would like to build deep learning that could process multi-label with a high-order strategy. That means all labels share the same learning process. It utilizes the relations information to further improve the performance.

### B. Transfer Learning

Transfer learning is a machine learning methodology that focuses on transferring knowledge across domains. The goal of transfer learning is to improve the performance of machine learning on target domains by transferring the knowledge learned from different but related source domains [16]. A common assumption in traditional machine learning is that training data and testing data have the same feature space and data distribution. When the distribution changes, most machine learning models need to be rebuilt. However, it is very difficult and expensive to recollect training data and rebuild the model in many real-world applications. Therefore, transfer learning between the source domain and target domain is quite desirable. With the development of deep learning, transfer learning has been a promising method in image analytics. Many studies have used the well-designed deep learning architecture for natural image datasets, such as IMAGENET, together with its pre-trained weights to fine-tune the model [11].

### C. Proposed Deep Learning Model

In this study, we aim to explore the role of a deep learning model for the multi-label classification of natural disasters. Specifically speaking, we would like to build a deep learning model using the high-order strategy [6] of multi-label learning to classify natural disasters. The high-order strategy addresses that connections exist among random subsets of labels. Instead of building several binary classification models independently, we build a deep convolutional neural network with multiple sigmoid functions as the activation function of the final output layer (Equation 1). We use the binary_crossentropy (Equation 2) as our loss function. This model could share the information in the multi-label training process. Utilizing label connections could help the performance of our model.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (1)$$

Z is the input of the sigmoid function, which is generated by the layer before the final out layer.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(p(y_i)) + (1-y_i)\, log(1-p(y_i))] \quad (2)$$

N is the number of classes. $y_i$ is the class indicator. $p(y_i)$is the probability of $y_i$ .

Fig.1 shows the framework work of our study. After getting the natural disaster image dataset, data cleaning work is required to get useful images with disaster scenes in the picture. Since each image contains 1 to 5 disaster scenes, we should transform the disaster labels into a one-hot encoding format. Each image has a one-hot label with 5 digits. Each digit corresponds to a disaster. 0 means the image does not contain this type of disaster while 1 means the image contains this type of disaster. And then, we resize all images to 150×150 pixels. Next, all image data will be fed to the deep convolutional neural network (DCNN) for disaster classification.

A DCNN is a feed-forward neural network commonly used in image analytics and comprises an input layer, hidden layers, and an output layer. The input is a tensor with a shape: inputs number × input height × input width × input channel. Hidden layers are convolutional layers followed by pooling layers. The convolutional layer contains several kernels to convolve the input and pass it to the next layer. A ReLU is used as the activation and could be expressed as follows:

$$f(x) = \max(0, x) \quad (3)$$

The pooling layer is a downsample method, which is used to reduce the number of parameters. Max pooling is often used in the DCNN model. The final layer is often a fully connected layer, which is used as a classifier to make the decision.

Since we focus on multi-label classification, we modify the deep convolution neural network (DCNN) by replacing the final output layer with multiple sigmoid functions. The number of sigmoid functions equals the number of class labels. To further speed up the training process and improve the classification performance, we use the transfer learning [16] method to strengthen our DCNN model. Transfer learning means transferring knowledge from one domain to another domain. In the deep learning research field, researchers usually use a pre-trained deep learning model to help the classification task. In this research, we use a pre-trained VGG16 [17] model to help the feature extraction. Fig. 2. shows the structure of our model.

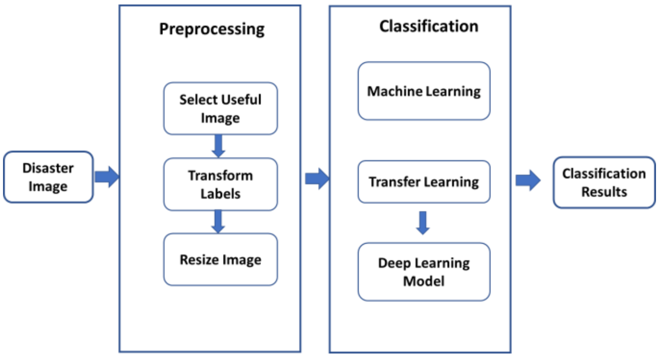


Fig.1. Research Framework

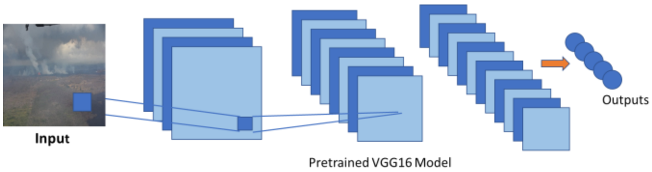


Fig.2. Proposed DCNN Model

*D. Disaster Image Dataset*

We extract the multi-label dataset from the LADI dataset [5] (https://github.com/mit-ll/ladi-overview#dataset-access). This dataset is collected by the Civil Air Patrol. There are various disaster scenes in this dataset, including flooding/water damage, landslide, road washout, rubble/debris, and smoke/fire. The low altitude, the oblique perspective of the imagery and the multi-label disaster make it distinctive compared with other datasets. The original dataset contains 39,981 images. Images with "no damage" or "damage(misc)" labels are excluded from the dataset since we focus on the specific natural disaster classification in this study. After that, we get 28,983 images. Each image has various disaster labels (from 1 to 5). There are 23,016 images with the "flooding/water damage" label. 1,973 images are labelled as "landslide". 3,006 images are named "road washout". 9535 images are labelled as "rubble/debris". 1598 images are named as "smoke/fire". Fig. 3 shows several samples of our image data.

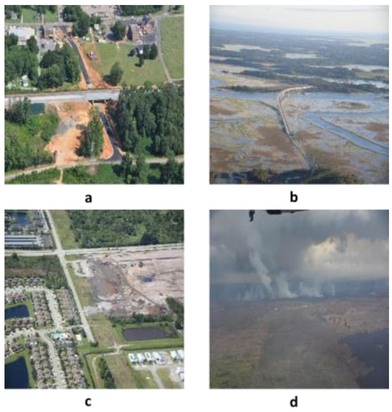

Fig.3. Sample of Disaster Images

Image a has two labels: "landslide" and "road washout"; Image b has two labels: "flooding/water damage" and "road washout"; Image c has two labels: "road washout" and "rubble/debris"; Image d has two labels: "rubble/debris" and "smoke/fire".

### *E. Evaluation Metrics*

To compare the performances of our proposed deep learning model with other machine learning models, five widely used metrics are computed to evaluate the overall performance. These metrics are accuracy, AUC, f1-score, precision, and recall [18].

## IV. EXPERIMENT RESULTS

Our experiments are carried out on a workstation with a 2.20 GHz Intel(R) Xeon(R) Silver 4210 CPU, 32 GB of memory, and an NVIDIA TITAN X GPU. The model was built and learned using TensorFlow, Keras, and scikit-learn. Those are three widely adopted open-source libraries for machine learning. The Adam optimizer with hyperparameters (learning rate=0.0001) is adopted to train the model [19]. Our deep learning model is trained for 300 epochs. An early stop with 15 patients is adopted. The batch size is 32. Each fold used 80% of the data as a training set, and 20% of the data was left for testing. We report the mean value of five folds as the final result.

### *A. Deep Learning vs. Traditional Machine Learning (Single Binary Classifier)*

We compared our DCNN model with traditional machine learning models, namely, decision tree (DT), and the multi-layer perception (MLP, two hidden layers, each layer with 32 neurons). Two sets of handcrafted image features are extracted: texture features (including contrast, dissimilarity, homogeneity, energy, correlation, and Angular Second Moment (ASM), all of them extracted from 0, ¼ π, ½ π, and ¾ π) (Mall et al. 2019) and Histogram of Oriented Gradient (HOG, 5 orientation bins, 25 pixels_per_cell, 5 cells for each block; Mizuno et al. 2012). These features were extracted using the scikit-image package. The transfer learning model used in this study was VGG16 pre-trained with IMAGENET. Except for our DCNN model, both decision tree and MLP use 5 binary classifiers to generate the multi-label results. The results are reported in Table I. We could find that our proposed DCNN model outperforms other models in all five metrics largely.

TABLE I. RESULTS OF TRADITIONAL MACHINE LEARNING METHODS (SINGLE BINARY CLASSIFIER)

| Model | Accuracy | AUC | F1-score | Precision | Recall |
|---|---|---|---|---|---|
| DT Texture | 0.7841 | 0.5427 | 0.3167 | 0.3122 | 0.3217 |
| DT HOG | 0.7634 | 0.5209 | 0.2957 | 0.2903 | 0.3033 |
| MLP Texture | 0.8476 | 0.5000 | 0.1772 | 0.1874 | 0.2001 |
| MLP HOG | 0.8506 | 0.5103 | 0.2260 | 0.2750 | 0.2317 |
| DCNN | **0.8881** | **0.7478** | **0.3460** | **0.4240** | **0.3377** |

### *B. Deep Learning vs. Traditional Machine Learning (Multi-label Classifier)*

We compared our DCNN model with the multi-layer perception (MLP, two hidden layers, each layer with 32 neurons). Instead of five binary classifiers, we built an MLP model with a multi-label classifier. The results are reported in Table III. We could find that our proposed DCNN model outperforms two MLP models in all five metrics largely.

TABLE III. RESULTS OF TRADITIONAL MACHINE LEARNING METHODS (MULTI-LABEL CLASSIFIER)

| Model | Accuracy | AUC | F1-score | Precision | Recall |
|---|---|---|---|---|---|
| MLP Texture | 0.8476 | 0.6049 | 0.1772 | 0.2655 | 0.2000 |
| MLP HOG | 0.8497 | 0.6104 | 0.2114 | 0.2718 | 0.2210 |
| DCNN | **0.8881** | **0.7478** | **0.3460** | **0.4240** | **0.3377** |

### *C. Deep Learning Models' Comparison (Multi-label Classifier)*

We compared three various deep models with the multi-layer perception: VGG16, ResNet50 [20], and InceptionV3[21]. The results are reported in Table IV. We could find that the difference among the three DCNN models is quite small. ResNet50 is the best one in F1-score, precision, and recall. VGG16 is the best one in accuracy and AUC.

TABLE IV. RESULTS OF VARIOUS DEEP LEARNING METHODS (MULTI-LABEL CLASSIFIER)

| Model | Accuracy | AUC | F1-score | Precision | Recall |
|---|---|---|---|---|---|
| ResNet50 | 0.8842 | 0.7338 | **0.3578** | **0.5668** | **0.3387** |
| InceptionV3 | 0.8858 | 0.7373 | 0.3457 | 0.5149 | 0.3280 |
| VGG16 | **0.8881** | **0.7478** | 0.3460 | 0.4240 | 0.3377 |

## V. DISCUSSION AND CONCLUSION

This study builds a DCNN model for multi-label classification of natural disaster classification. We train and test the model using a high-quality low altitude, oblique perspective disaster dataset. We compare the DCNN model with traditional machine learning with a first-order strategy. We find our DCNN model with transfer learning performs better than other models. We also compare our DCNN model with traditional machine learning with the same high-order strategy. Our DCNN model also outperforms other models. Finally, we explore the effects of different DCNN models on multi-label classification. We find their difference is quite small.

Our study contributes to multi-label classification in the context of natural disaster response and management. Social media data in the previous study may not very accurate and reliable for decision making. Utilizing the low altitude, oblique perspective disaster dataset collected by a professional organization, we could provide more support to disaster response. Besides, our DCNN model could cope with multi-label classification, which has a wide application scenario because multiple disasters usually happen at the same time in reality. It could help the decision-making in disaster response and management. Finally, this study also demonstrates how to apply image processing, deep learning, and transfer learning methods in a specific context. Although image data is more complex than numeric and text data, our multi-label image classification framework provides some insights for researchers and practitioners.

Our study also has several limitations. First, we only use image data for multi-label classification. Multi-modal data could provide more useful information for decision-making. Second, those image data are collected from the USA. Whether our model performs well in another environment is not clear. However, deep learning has made great progress in image classification in many fields. And we also have transfer learning, which could also help the extension in another dataset. We think that our model can transfer to another dataset. Therefore, we will continue our research in this field in the future. We will explore disaster classification using multi-modal data in the future.

## ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2019YFC1509402).

## REFERENCE

[1] H. Hao and Y. Wang, "Leveraging multimodal social media data for rapid disaster damage assessment," *International Journal of Disaster Risk Reduction*, vol. 51, p. 101760, Dec. 2020, doi: 10.1016/j.ijdrr.2020.101760.

[2] W. Gao, L. Li, X. Zhu, and Y. Wang, "Detecting Disaster-Related Tweets Via Multimodal Adversarial Neural Network," *IEEE MultiMedia*, vol. 27, no. 4, pp. 28–37, 2020, doi: 10.1109/MMUL.2020.3012675.

[3] L. (Lucy) Yan and A. Pedraza Martinez, "Social Media for Disaster Management: A Study of Hurricane Sandy," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2716217, Sep. 2017. doi: 10.2139/ssrn.2716217.

[4] M. Imran, F. Ofli, D. Caragea, and A. Torralba, "Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions," *Information Processing & Management*, vol. 57, no. 5, p. 102261, Sep. 2020, doi: 10.1016/j.ipm.2020.102261.

[5] J. Liu, D. Strohschein, S. Samsi, and A. Weinert, "Large Scale Organization and Inference of an Imagery Dataset for Public Safety," in *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, Waltham, MA, USA, Sep. 2019, pp. 1–6. doi: 10.1109/HPEC.2019.8916437.

[6] M.-L. Zhang and Z.-H. Zhou, "A Review on Multi-Label Learning Algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014, doi: 10.1109/TKDE.2013.39.

[7] Ying Lin, "10 Twitter Statistics Every Marketer Should Know in 2020 [Infographic]," *Marketing Your Store*, Nov. 30, 2019. https://www.oberlo.com/blog/twitter-statistics (accessed Aug. 24, 2020).

[8] F. Alam, T. Alam, M. A. Hasan, A. Hasnat, M. Imran, and F. Ofli, "MEDIC: A Multi-Task Learning Dataset for Disaster Image Classification," *arXiv:2108.12828 [cs]*, Dec. 2021, Accessed: Feb. 20, 2022. [Online]. Available: http://arxiv.org/abs/2108.12828

[9] A. Asif *et al.*, "Automatic analysis of social media images to identify disaster type and infer appropriate emergency response," *Journal of Big Data*, vol. 8, no. 1, p. 83, Jun. 2021, doi: 10.1186/s40537-021-00471-5.

[10] Y. Han, P. Cui, Y. Zhang, R. Zhou, S. Yang, and J. Wang, "Remote Sensing Sea Ice Image Classification Based on Multilevel Feature Fusion and Residual Network," *Mathematical Problems in Engineering*, vol. 2021, p. e9928351, Sep. 2021, doi: 10.1155/2021/9928351.

[11] I. Bouchard, M.-È. Rancourt, D. Aloise, and F. Kalaitzis, "On Transfer Learning for Building Damage Assessment from Satellite Imagery in Emergency Contexts," *Remote Sensing*, vol. 14, no. 11, Art. no. 11, Jan. 2022, doi: 10.3390/rs14112532.

[12] X. Wang, D. He, F. Feng, J. A. Ashton-Miller, J. O. L. DeLancey, and J. Luo, "Multi-label classification of pelvic organ prolapse using stress magnetic resonance imaging with deep learning," *Int Urogynecol J*, Jan. 2022, doi: 10.1007/s00192-021-05064-7.

[13] X. Pan *et al.*, "Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning," *Graefes Arch Clin Exp Ophthalmol*, vol. 258, no. 4, pp. 779–785, Apr. 2020, doi: 10.1007/s00417-019-04575-w.

[14] K. V. Priya and J. D. Peter, "A federated approach for detecting the chest diseases using DenseNet for multi-label classification," *Complex Intell. Syst.*, vol. 8, no. 4, pp. 3121–3129, Aug. 2022, doi: 10.1007/s40747-021-00474-y.

[15] R. Yuan, J. Ni, and Q. Zhou, "Generating Multimedia Storyline for Effective Disaster Information Awareness," *IEEE Access*, vol. 7, pp. 47401–47410, 2019, doi: 10.1109/ACCESS.2019.2907164.

[16] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: May 11, 2022. [Online]. Available: http://arxiv.org/abs/1409.1556

[18] N. Chaudhuri and I. Bose, "Exploring the role of deep neural networks for post-disaster decision support," *Decision Support Systems*, vol. 130, p. 113234, Mar. 2020, doi: 10.1016/j.dss.2019.113234.

[19] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. doi: 10.48550/arXiv.1412.6980.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision." arXiv, Dec. 11, 2015. doi: 10.48550/arXiv.1512.00567.